Incremental Causal Reasoning

Thomas Dean and **Mark Boddy**[1]
Department of Computer Science
Brown University
Providence, RI 02912

Abstract

Causal reasoning comprises a large portion of the inference performed by automatic planners. In this paper, we consider a class of inference systems that are said to be *predictive* in that they derive certain causal consequences of a base set of premises corresponding to a set of events and constraints on their occurrence. The inference system is provided with a set of rules, referred to as a *causal theory*, that specifies, with some limited accuracy, the cause and effect relationships between objects and processes in a given domain. As modifications are made to the base set of premises, the inference system is responsible for accounting for all and only those inferences licensed by the premises and current causal theory. Unfortunately, the general decision problem for nontrivial causal theories involving partially ordered events is *NP-complete*. As an alternative to a complete but potentially exponential-time inference procedure, we describe a limited-inference polynomial-time algorithm capable of dealing with partially ordered events. This algorithm generates a useful subset of those inferences that will be true in all total orders consistent with some specified partial order. The algorithm is incremental and, while it is not complete, it is provably sound.

I. Introduction

We are concerned with the process of incrementally constructing *nonlinear* plans (i.e., plans represented as sets of actions whose order is only partially specified). A significant part of this process involves some means for predicting the consequences of actions and using these consequences to verify whether or not a given partially constructed plan is likely to succeed. Of course, if by "likely to succeed" we mean that the choices made thus far in constructing the partial plan will not require further revision, then it is obvious that this verification step subsumes the entire process of planning. Usually, by "likely to succeed" we mean something like: given a partially ordered set of tasks and their intended effects, make sure that there is at least one total ordering consistent with the initial partial order such that all of the tasks have their intended effects. At first blush, determining whether a given partially constructed plan satisfies this criterion appears to be a significantly easier problem than the general planning problem. Unfortunately, if the language used to represent plans, tasks, and their effects is sufficiently expressive and we use asymptotic complexity as our measure of difficulty, the problem faced by the temporal reasoning component is just as difficult as the general planning problem. This shouldn't surprise anyone, but neither should it discourage anyone from employing classical planning techniques. It does indicate, however, that we have some way to go in understanding the expressive and computational requirements for effective temporal reasoning systems.

A theory for reasoning about the effects of actions (or, more generally, the consequences of events) we refer to as a *causal theory*. We will describe a language for constructing causal theories that is capable of representing indirect effects and actions whose effects depend upon the situation in which the actions occur. We will consider two algorithms for reasoning about such causal theories. These algorithms are polynomial-time, incremental, and insensitive to the order in which facts are added to or deleted from the data base. We show that one algorithm is complete for causal theories in which the events are totally ordered, but is potentially inconsistent in cases where the events are not totally ordered. The general problem of reasoning about the effects of actions that are partially ordered and whose effects depend upon the situation in which the actions occur has been shown to be *NP-hard* [1]. As an alternative to a complete but potentially exponential-time decision procedure, we provide a partial decision procedure that is provably sound. What this means for a planner is that the procedure is guaranteed not to mislead the planner into committing to a plan that is provably impossible given what is currently known. If the decision procedure answers yes, then the condition in question is guaranteed to hold in every totally ordered extension of the current partial order; if the decision procedure answers no, there is a chance that the condition holds in every total order, but to determine this with certainty might require an exponential amount of time or space.

II. Temporal Data Base Management

A temporal data base management system (TDBMS) is used to keep track of what is known about the order, duration,

[1]This work was supported in part by the National Science Foundation under grant IRI-8612644 and by an IBM faculty development award.

and time of occurrence of a set of events and their consequences. In the rest of this paper, we will be concerned with a particular type of TDBMS called a *time map management system* or TMM. [3]. In the TMM, the classical data base assertion is replaced by the notion of *time token* corresponding to a particular interval of time during which a general *type* of occurrence (a fact or event) is said to be true. For any given fact or event type, the data base (or time map) will typically include many tokens of that type.

The user of the TMM can specify information concerning events that have been observed or are assumed inevitable and information in the form of general rules that are believed to govern the physics of a particular domain. The user can also specify certain conditional beliefs. If the user explicitly states the conditions for believing in certain propositions, the TMM can ensure that those propositions (and their consequences) are present in the data base just in case the conditions are met. This is achieved through the use of *data dependencies* [7]. In the TMM, the primary forms of data dependency (in addition to those common in static situations) are concerned with some fact being true at a point in time or throughout an interval. In addition, there is a nonmonotonic form of temporal data dependency concerned with it being *consistent* to believe that a fact is not true at a point in time or during any part of an interval. These forms of temporal data dependency are handled in the TMM using the mechanism of *temporal reason maintenance* [3]. Language constructs are supplied in the TMM that allow an application program to query the data base in order to establish certain antecedent conditions (including temporal conditions) and then, on the basis of these conditions, to assert consequent predictions. These predictions remain valid just in case the antecedent conditions continue to hold.

Perhaps the most important and most often overlooked characteristic of a temporal reasoning system is its ability to handle incomplete information of the sort one invariably encounters in realistic applications. For example, we seldom know the exact duration or time of occurrence of most events. Moreover, for those durations and offsets we do know, they are seldom with respect to a global frame of reference such as a clock or calendar. In the TMM, every point is a frame of reference, and it is possible to constrain the distance between any two points simply by specifying bounds, $\langle low, high \rangle$, on the distance in time separating the two points. By allowing bounds to be both numeric and symbolic, the same framework supports both qualitative and quantitative relationships.

Another important aspect of reasoning with incomplete information has to do with the default character of temporal inference. In general, it is difficult to predict in advance how long a fact made true will persist. It would be convenient to leave it up to the system to decide how long facts persist based upon the simple default rule [9] that a fact made true continues to be so until something serves to make it false. This is exactly what the TMM does. The term *persistence* is used to refer to an interval corresponding to a particular (type of) fact becoming true and remaining so for some length of time. A fact is determined to be true throughout an interval I just in case there is a persistence that begins before the beginning of I and it can't be shown that the persistence ends before the end of I. Before we continue our discussion it will help to introduce some notation.

Relations. Let Π be the set of points corresponding to the begin and end of events in a particular temporal data base. We define a function DIST to denote the best known bounds on the distance in time separating two points. Given $\pi_1, \pi_2 \in \Pi$ such that $\text{DIST}(\pi_1, \pi_2) = \langle low, high \rangle$, we have:

- $\pi_1 \prec \pi_2 \Leftrightarrow low \geq \epsilon^2$ (π_1 is ordered before π_2)
- $\pi_1 \equiv \pi_2 \Leftrightarrow \langle low, high \rangle = \langle 0, 0 \rangle$ (π_1 is coincident with π_2)
- $\pi_1 \preceq \pi_2 \Leftrightarrow (\pi_1 \prec \pi_2) \vee (\pi_1 \equiv \pi_2)$ (π_1 precedes or is coincident with π_2)
- $\pi_1 \prec_M \pi_2 \Leftrightarrow high \geq \epsilon$ (π_1 possibly precedes π_2)
- $\pi_1 \preceq_M \pi_2 \Leftrightarrow high \geq 0$ (π_1 possibly precedes or is coincident with π_2)

Tokens. We denote a set of *time tokens* $\mathsf{T} = \{t_0, t_1, \ldots t_n\}$ for referring to intervals of time during which certain events occur or certain facts are known to become true and remain so for some period of time. The latter correspond to what we have been calling persistences. For a given token t:

- $\text{BEGIN}(t)$, $\text{END}(t) \in \Pi$.
- $\text{STATUS}(t) \in \{\text{IN}, \text{OUT}\}$, determined by whether the token is warranted (`IN`) or not (`OUT`) by the current premises and causal theory.
- $\text{TYPE}(t) = P$ where P is an atomic predicate calculus formula with no variables.
- $\text{DURATION}(t) = \text{DIST}(\text{BEGIN}(t), \text{END}(t))$

Types. As defined above, the type of an individual token is an atomic formula with no variables (e.g., `(on block14 table42)`). In general, any atomic formula, including those containing variables, can be used to specify a type. In describing the user interface, universally quantified variables are notated `?variable-name`, the scope of the variable being the entire formula in which it is contained (e.g., `(on ?x ?y)`). In describing the behavior of the inference system, we will use variables of the form t_P to quantify over tokens of type P (i.e., $\forall t_P \in \mathsf{T}\ \text{TYPE}(t_P) = P$).

As we will see in the next section, the TMM allows a user to specify rules (referred to collectively as a *causal theory*) for inferring additional consequences of the data (referred to as the *set of basic facts* and notated B). B consists of a set of time tokens and a set of constraints on

[2] The symbol ϵ is meant to denote an infinitesimal: a number greater than 0 and smaller than any positive number.

R1: (project (and $P_1 \ldots P_n$
(M (not (and $Q_1 \ldots Q_m$))))
E R)

R2: (project (and $P_1 \ldots P_n$)
E R)

R3: (disable (and $Q_1 \ldots Q_m$)
(ab R2))

R4: (disable (and $R_1 \ldots R_o$)
(ab R3))

Figure 1: Hierarchically arranged projection and disabling rules

the amount of time separating pairs of points corresponding to the begin and end of time tokens. Generally, the causal theory remains fixed for a specific application, and a program interacts with the TMM by adding and removing items from B, and by generating queries. A query consists of a predicate calculus formula corresponding to a question of the form "Could some fact P be true over a particular interval I?" An affirmative answer returned by the TMM in response to such a query will include a set of assumptions necessary for concluding that the fact is indeed true. Any assertions made on the basis of the answer to such a query are made to depend upon these assumptions.

The state of a temporal data base is completely defined by a temporal constraint graph (TCG), consisting of the points in Π and constraints between them, and a causal dependency graph (CDG), consisting of dependency structures corresponding to the application of causal rules in deriving new tokens. The TCG and CDG are incrementally modified to reflect changes in the set B.

III. Causal Theories

In the TMM, a causal theory is simply a collection of rules, called *projection rules*, that are used to specify the behavior of processes. In the following rule, $P_1 \ldots P_n$, $Q_1 \ldots Q_m$, E, and R designate types, and *delay* and *duration* designate constraints (e.g., $\langle \epsilon, \infty \rangle$). In:

(project (and $P_1 \ldots P_n$
(M (not (and $Q_1 \ldots Q_m$))))
E *delay* R *duration*)

$P_1 \ldots P_n$ and $Q_1 \ldots Q_m$ are referred to as *antecedent conditions*, E is the type of the *triggering event*, and R refers to the type of the *consequent prediction*. The above projection rule states that, if an event of type E occurs corresponding to the token t_E and $P_1 \ldots P_n$ are believed to be true at the outset[3] of t_E and it is consistent to believe that the conjunction of $Q_1 \ldots Q_m$ is not true at the outset of t_E, then, after an interval of time following the end of t_E determined by *delay*, R will become true and remain so for a period of time constrained by *duration* (if *delay* and *duration* are not specified, they default to $\langle 0, 0 \rangle$ and $\langle \epsilon, \infty \rangle$, respectively). In the following, we will be considering a restricted form of causal theory, called a *type 1* theory, such that the *delay* always specifies a positive offset (causes always precede their effects).

We also allow the user to specify rules that serve to *disable* other rules [11]. Figure 1 shows a standard projection rule R1 and a pair of projection and disabling rules R2 and R3 that replace R1. The rule R3 is further conditioned by the rule R4. Assuming just the rules R2, R3, and R4, any application of R2 with respect to a particular token t of type E is said to be abnormal with regard to t just in case $Q_1 \ldots Q_m$ hold at the outset of t and it is consistent to believe that R3 is not abnormal with regard to t. The nonmonotonic behavior of type 1 causal theories is specified entirely in terms of disabling rules and the default rule of persistence (see Section II.). In addition to their usefulness for handling various forms of incomplete information, disabling rules make it possible to reason about the consequences of simultaneous actions. The reader interested in a more detailed treatment of causal theories may refer to one of [5] or [11]. Throughout the rest of this paper we will consider causal theories without disabling rules and consisting solely of *simple projection rules*[4]. The following represents the general form of a simple projection rule:

(project (and $P_1 \ldots P_n$) E *delay* R *duration*)

In order to support temporal reasoning, there has to be some method or *decision procedure* for drawing appropriate conclusions from a set of basic facts and a given causal theory. In the TMM, such a procedure is used to generate new time tokens, update the status of existing tokens, and facilitate query processing by determining the truth of facts over specified intervals of time. As far as we are concerned, an inference procedure is fully specified by a criterion for inferring consequent effects from antecedent causes via causal rules, a method for actually applying that criterion (an update algorithm), and a criterion for determining if a fact is true throughout some interval. Figure 2 shows a criterion for inferring consequent effects which we refer to as *weak projection*. The criterion is specified as a rule schema (implicitly) quantified over simple projection rules. Figure 3 shows a criterion for determining if a fact is true throughout an interval which we refer to as *weak true throughout*. The criterion is specified in terms of a

[3] An alternative formulation described in [6] states that the antecedent conditions of a projection rule must be true *throughout* the trigger event rather than true just at the outset. Both formulations are supported in the TMM, though we will only be discussing the true-at-the-outset formulation in this paper.

[4] All of the results mentioned in this paper extend to full type 1 theories (see [4]).

$$\begin{array}{l}
\forall t_E \in \mathsf{T} \\
\quad ((\textsc{status}(t_E) = \texttt{IN}) \wedge \\
\quad (\exists t_{P_1} \ldots t_{P_n} \in \mathsf{T} \\
\quad\quad (\forall\ 1 \leq i \leq n\ (\textsc{status}(t_{P_i}) = \texttt{IN}) \wedge \\
\quad\quad\quad (\textsc{begin}(t_{P_i}) \preceq \textsc{begin}(t_E)) \wedge \\
\quad\quad\quad (\forall t_{\neg P_i} \in \mathsf{T} \\
\quad\quad\quad\quad (\textsc{status}(t_{\neg P_i}) = \texttt{OUT}) \vee \\
\quad\quad\quad\quad (\textsc{begin}(t_{\neg P_i}) \prec_M \textsc{begin}(t_{P_i})) \vee \\
\quad\quad\quad\quad (\textsc{begin}(t_E) \prec_M \textsc{begin}(t_{\neg P_i}))\ )))) \\
\quad \Rightarrow\ \exists t_R \in \mathsf{T} \\
\quad\quad (\textsc{status}(t_R) = \texttt{IN}) \wedge \\
\quad\quad (\textsc{dist}(\textsc{end}(t_E), \textsc{begin}(t_R)) \subseteq \mathit{delay}) \wedge \\
\quad\quad (\textsc{dist}(\textsc{begin}(t_R), \textsc{end}(t_R)) \subseteq \mathit{duration})
\end{array}$$

Figure 2: Weak projection

$$\begin{array}{l}
\forall \pi_1 \pi_2 \in \Pi \\
\quad \exists t_P \in \mathsf{T} \\
\quad\quad (\textsc{status}(t_P) = \texttt{IN}) \wedge \\
\quad\quad (\textsc{begin}(t_P) \preceq \pi_1) \wedge \\
\quad\quad (\forall t_{\neg P} \in \mathsf{T} \\
\quad\quad\quad (\textsc{status}(t_{\neg P}) = \texttt{OUT}) \vee \\
\quad\quad\quad (\textsc{begin}(t_{\neg P}) \prec_M \textsc{begin}(t_P)) \vee \\
\quad\quad\quad (\pi_2 \prec_M \textsc{begin}(t_{\neg P_i}))\ )))) \\
\quad \Leftrightarrow\ TT(P, \pi_1, \pi_2)
\end{array}$$

Figure 3: Weak true throughout

definition of the true throughout predicate TT. The inference procedure (referred to as *naive projection*) consisting of weak projection, weak true throughout, and a simple update algorithm for applying weak projection by sweeping forward in time was used in one of the early versions of the TMM. In the following section, we will consider some of the properties of naive projection.

IV. Completeness and Consistency

In order to satisfy ourselves concerning the behavior of an inference system, we need a precise account of what the conclusions computed by that system *mean*. Such an account should enable us to judge whether or not an inference system has come up with the *right* set of conclusions. The question we need to ask is: What are the intended models of a set of basic facts and a causal theory?

As far as we are concerned, a *model* consists of an assignment of true or false to a particular set of propositions concerning facts spanning intervals of time. Theories about the real world are invariably underconstrained, and a set of basic facts together with a causal theory will generally have many models. We will simplify our analysis by partitioning models into various equivalence classes. The primary source of ambiguity in the TMM arises from the fact that the set of constraints seldom determines a total ordering of the tokens in T. Given that most inferences depend only upon what is true during intervals defined by points corresponding to the begin and end of tokens in T, all that we are really interested in are the classes of models corresponding to the different total orderings consistent with the initial set of constraints. For each total ordering we can identify a unique set of tokens that intuitively should be IN given a particular causal theory.

We start with a set of basic facts B, consisting of a set of tokens $\mathsf{T_B}$ and a set of constraints $\mathsf{C_B}$. The constraints in $\mathsf{C_B}$ determine a partial order on the begin and end of tokens in $\mathsf{T_B}$. For a particular B, there may be a number of total orderings consistent with the constraints in $\mathsf{C_B}$.

For a given B, a fixed causal theory, and a criterion for inferring consequent effects from antecedent causes (e.g., weak projection), the TMM generates a set of tokens T and a temporal constraint graph (TCG). Given T and the TCG, there are a finite number of statements of the form $TT(P, \pi_1, \pi_2)$ that are determined as true by the TMM using a particular true throughout criterion (e.g., weak true throughout). The criterion for inferring consequent effects must be applied in a systematic way (essentially using the ordering information to perform a sweep forward in time) to yield results in keeping with our intuitions about causality. The strategy built into the TMM for applying the criterion of weak projection with respect to specific tokens and updating the status of tokens already in T makes use of the intuition that you can't know the effects of a particular event e until you know the consequences of those events preceding e. It should be fairly easy to convince yourself that, in cases in which $\mathsf{C_B}$ precisely constrains the order of the tokens in $\mathsf{T_B}$, the TMM, using weak projection, generates a set T and a TCG such that the statements of the form $TT(P, \pi_1, \pi_2)$ determined true by the weak true throughout criterion are exactly the ones that we want. We will make use of this to define a working notion of model.

Given some B together with a fixed causal theory, for each total ordering consistent with $\mathsf{C_B}$, we will say that the set of statements of the form $TT(P, \pi_1, \pi_2)$ that are true using weak true throughout and weak projection is a model of B and the underlying causal theory. This set can be thought of as specifying an assignment to just those statements concerned with facts being true over intervals. Actually, the assignment designates a class of models, but we will neglect this to simplify our discussion. We will say that a particular inference procedure is *complete* for a class of causal theories, if for any set of basic facts and causal theory in that class, the statements of the form $TT(P, \pi_1, \pi_2)$ warranted by the inference procedure include at least those that are true in all models. Similarly, we will say that an inference procedure is *sound* for a class of causal theories, if for any set of basic facts and causal theory in that class, each statement $TT(P, \pi_1, \pi_2)$ warranted by the inference procedure is true in all models.

Given the preceding definitions, it is easy to show that the TMM, using naive projection, is complete and sound for type 1 causal theories, assuming that the tokens in T are totally ordered [4].

$$
\begin{array}{l}
\forall t \in \mathsf{T}_{\mathsf{B}} \\
\quad \mathit{STRONGLY\text{-}PROTECTED}(t)
\end{array}
$$

$$
\begin{array}{l}
\forall t_E \in \mathsf{T} \\
\quad (\mathit{STRONGLY\text{-}PROTECTED}(t_E) \wedge \\
\quad (\exists t_{P_1} \ldots t_{P_n} \in \mathsf{T} \\
\quad (\forall\ 1 \leq i \leq n\ \mathit{STRONGLY\text{-}PROTECTED}(t_{P_i}) \wedge \\
\qquad (\mathrm{BEGIN}(t_{P_i}) \preceq \mathrm{BEGIN}(t_E)) \wedge \\
\qquad (\forall t_{\neg P_i} \in \mathsf{T} \\
\qquad\quad (\mathrm{STATUS}(t_{\neg P_i}) = \mathtt{OUT}) \vee \\
\qquad\quad (\mathrm{BEGIN}(t_{\neg P_i}) \prec \mathrm{BEGIN}(t_{P_i})) \vee \\
\qquad\quad (\mathrm{BEGIN}(t_E) \prec \mathrm{BEGIN}(t_{\neg P_i}))\ )))) \\
\Rightarrow \exists t_R \in \mathsf{T} \\
\qquad \mathit{STRONGLY\text{-}PROTECTED}(t_R) \wedge \\
\qquad (\mathrm{DIST}(\mathrm{END}(t_E), \mathrm{BEGIN}(t_R)) \subseteq \mathit{delay}) \wedge \\
\qquad (\mathrm{DIST}(\mathrm{BEGIN}(t_R), \mathrm{END}(t_R)) \subseteq \mathit{duration})
\end{array}
$$

Figure 4: Strongly protected tokens

$$
\begin{array}{l}
\forall t_E \in \mathsf{T} \\
\quad ((\mathrm{STATUS}(t_E) = \mathtt{IN}) \wedge \\
\quad (\exists t_{P_1} \ldots t_{P_n} \in \mathsf{T} \\
\quad (\forall\ 1 \leq i \leq n\ (\mathrm{STATUS}(t_{P_i}) = \mathtt{IN}) \wedge \\
\qquad (\mathrm{BEGIN}(t_{P_i}) \preceq_M \mathrm{BEGIN}(t_E)) \wedge \\
\qquad (\forall t_{\neg P_i} \in \mathsf{T} \\
\qquad\quad \neg \mathit{STRONGLY\text{-}PROTECTED}(t_{\neg P_i}) \vee \\
\qquad\quad (\mathrm{BEGIN}(t_{\neg P_i}) \prec_M \mathrm{BEGIN}(t_{P_i})) \vee \\
\qquad\quad (\mathrm{BEGIN}(t_E) \prec_M \mathrm{BEGIN}(t_{\neg P_i}))\ )))) \\
\Rightarrow \exists t_R \in \mathsf{T} \\
\qquad (\mathrm{STATUS}(t_R) = \mathtt{IN}) \wedge \\
\qquad (\mathrm{DIST}(\mathrm{END}(t_E), \mathrm{BEGIN}(t_R)) \subseteq \mathit{delay}) \wedge \\
\qquad (\mathrm{DIST}(\mathrm{BEGIN}(t_R), \mathrm{END}(t_R)) \subseteq \mathit{duration})
\end{array}
$$

Figure 5: Improbably weak projection

In situations where the set of basic facts does not determine a total order, it is easy to show that the TMM, using naive projection, can end up in a state with IN tokens that allow one to conclude statements of the form $TT(P, \pi_1, \pi_2)$ that are not true in *any* totally ordered extension. In [4], we prove that the problem of determining if $TT(P, \pi_1, \pi_2)$ is true for a type 1 causal theory, with or without disabling rules, is *NP-complete*.

In the rest of this paper, we abandon the quest for complete inference procedures and concern ourselves with procedures that are sound. To improve the chances of the TMM warranting only valid statements of the form $TT(P, \pi_1, \pi_2)$ the first thing we will do is strengthen the criterion for belief in a given token. The axioms in Figure 4 determine a set of tokens that are said to be strongly protected. If the set of constraints determines a total ordering, then the set of strongly protected tokens is identical to the set of tokens that are IN, but generally the former is a subset of the latter. Next, we provide a criterion for generating consequent predictions that takes into account

$$
\begin{array}{l}
\forall \pi_1 \pi_2 \in \Pi \\
\quad \exists\ t_P \in \mathsf{T} \\
\qquad \mathit{STRONGLY\text{-}PROTECTED}(t_P) \wedge \\
\qquad (\mathrm{BEGIN}(t_P) \preceq \pi_1) \wedge \\
\qquad (\forall t_{\neg P} \in \mathsf{T} \\
\qquad\quad (\mathrm{STATUS}(t_{\neg P}) = \mathtt{OUT}) \vee \\
\qquad\quad (\mathrm{BEGIN}(t_{\neg P}) \prec \mathrm{BEGIN}(t_P)) \vee \\
\qquad\quad (\pi_2 \prec \mathrm{BEGIN}(t_{\neg P}))\ ) \\
\quad \Leftrightarrow\ TT(P, \pi_1, \pi_2)
\end{array}
$$

Figure 6: Strong true throughout

every consequence that might be true in any total order, called *improbably weak projection.* This criterion is shown in Figure 5. And, finally, we provide a criterion for true throughout that succeeds only if the corresponding formula will be true in all total orders consistent with the current set of constraints (see Figure 6).

There is a simple decision procedure for generating all consequences and computing the set of strongly protected tokens. Let $\mathsf{T}_0 = \mathsf{T}_{\mathsf{B}}$, and initially assume that no tokens are strongly protected. Let $i = 0$. To compute the consequences of T_i, generate the consequent tokens of each token in T_i using the criterion of improbably weak projection. Let T_{i+1} be the union of T_i and its consequences. Continue to compute new consequent tokens in this manner, incrementing i as needed until $\mathsf{T}_i = \mathsf{T}_{i+1}$. Set $\mathsf{T} = \mathsf{T}_i$. At this point, perform a sweep forward in time (relative to the current partial order) determining for each token in T whether or not it is strongly protected and the status, IN or OUT, of each its consequents. In [4], we prove that this decision procedure is sound for a partially ordered set of tokens, and sound and complete for a totally ordered set.

In [4], we describe an incremental update algorithm that has the same soundness and completeness properties as the algorithm described above. This incremental algorithm is such that small changes in B generally result in small amounts of computation. For causal theories in which the consequent predictions of causal rules all correspond to persistences, the worst-case behavior of the incremental algorithm is polynomial in the size of B and the causal theory. If we allow causal rules to generate new tokens corresponding to the occurrence of triggering events, it is easy to construct examples in which T grows without bound. Generally, however, even those causal theories that generate new triggering events turn out to be well behaved. In a planning system, the incremental algorithm can be used as part of a strategy for coping with complexity; if a query succeeds, the answer can be assured to be true in all totally ordered extensions. If, on the other hand, a query fails and the truth or falsity of the query is critical, the system can choose to expend additional effort in processing the query. In [4] we describe some additional techniques that can be used to improve the accuracy of our decision procedure without sacrificing its performance (e.g., a simple examination of the tokens in T can serve to guarantee the failure of certain queries).

V. Delayed-Commitment Planning

Nonlinear planning [2] has long been considered to have distinct advantages over linear planning systems such as STRIPS [8] and its descendents. One supposed advantage [10] has to do with the idea that, by delaying commitment to the order in which "independent" actions are to be performed, a planner can avoid unnecessary backtracking. Linear planners are often forced to make arbitrary commitments regarding the order in which actions are to be carried out. Such arbitrary orderings often fail to lead to a solution and have to be reversed. By ordering only actions known to interact with one another (i.e., actions whose outcomes depend upon the order in which the actions are executed) the expectation was that nonlinear planners would avoid a lot of unnecessary work.

The problem in getting this sort of delayed-commitment planning to work is that it is often difficult to determine if two actions actually are independent. This is especially so if we are considering a representation of actions sufficiently powerful to represent actions whose effects depend upon context. In order to determine whether or not two actions are independent, it is necessary to determine what the effects of those actions are. Unfortunately, in order to determine the effects of a given action it is necessary to determine what is true prior to that action being executed, and this in turn requires that we know the effects of those actions that precede that action. In general there is no way to determine whether or not two actions are independent without actually considering all of the possible total orderings involving those two actions.

Planning depends upon the ability to predict the consequences of acting. Past planning systems capable of reasoning about partial orders (i.e., nonlinear planners) have either employed weak (and often unsound) methods for performing predictive inference or they have sought to delay prediction until the conditions immediately preceding an action are known with certainty. Delaying predictive inference can serve to avoid inconsistency, but it can also result in extensive backtracking in those very situations that nonlinear planners were designed to handle efficiently.

It is our contention that delayed-commitment planning is of dubious utility. However, the idea of delayed-commitment planning is not the only reason for building planners capable of reasoning about partially ordered events. Most events will not be under a planner's control and more often than not it will be difficult if not impossible to determine the order of all events with absolute certainty. Reasoning about partially ordered events is likely to play a significant role in future planners.

VI. Conclusions

This paper is concerned with computational approaches to reasoning about time and causality, particularly in domains involving partial orders and incomplete information. We have described a class of causal theories capable of representing conditional effects and the effects of simultaneous actions. We have described a decision procedure for generating predictions warranted by such causal theories. The decision procedure is provably sound and the resulting conclusions are guaranteed consistent if the underlying causal theory is consistent. If the events turn out to be totally ordered, the procedure is complete as well as sound.

Acknowledgments

This work was instigated in part by an offhand remark made by Kurt Konolige while the first author was visiting SRI International. The different approaches to planning of David Chapman and David Wilkins were both influential in directing our research.

References

1. Chapman, David, *Planning for Conjunctive Goals,* Technical Report AI-TR-802, MIT AI Laboratory, 1985.
2. Charniak, Eugene and McDermott, Drew V., *Introduction to Artificial Intelligence* (Addison-Wesley Publishing Co., 1985).
3. Dean, Thomas, and McDermott, Drew V., Temporal Data Base Management, *Artificial Intelligence* **32** (1987).
4. Dean, Thomas, and Boddy, Mark, *Incremental Causal Reasoning*, Technical Report CS-87-01, Brown University Department of Computer Science, 1987.
5. Dean, Thomas, An Approach to Reasoning About the Effects of Actions for Automated Planning Systems, *Annals of Operations Research* (1987).
6. Dean, Thomas, *Temporal Imagery: An Approach to Reasoning about Time for Planning and Problem Solving,* Technical Report 433, Yale University Computer Science Department, 1985.
7. Doyle, Jon, A truth maintenance system, *Artificial Intelligence* **12** (1979) 231–272.
8. Fikes, Richard and Nilsson, Nils J., STRIPS: A new approach to the application of theorem proving to problem solving, *Artificial Intelligence* **2** (1971) 189–208.
9. Reiter, Raymond, A Logic for Default Reasoning, *Artificial Intelligence* **13** (1980).
10. Sacerdoti, Earl, *A Structure for Plans and Behavior* (American Elsevier Publishing Company, Inc., 1977).
11. Shoham, Yoav, *Chronological Ignorance: Time, Knowledge, Nonmonotonicity and Causation, Proceedings AAAI-86*, Philadelphia, PA, AAAI, 1986.